Year: 2015

Diagnosis of esophageal motility disorders: esophageal pressure topography vs. conventional line tracing

Carlson, Dustin A ; Ravi, Karthik ; Kahrilas, Peter J ; Gyawali, C Prakash ; Bredenoord, Arjan J ; Castell, Donald O ; Spechler, Stuart J ; Halland, Magnus ; Kanuri, Navya ; Katzka, David A ; Leggett, Cadman L ; Roman, Sabine ; Saenz, Jose B ; Sayuk, Gregory S ; Wong, Alan C ; Yadlapati, Rena ; Ciolino, Jody D ; Fox, Mark R ; Pandolfino, John E

Abstract: OBJECTIVES: Enhanced characterization of esophageal peristaltic and sphincter function provided by esophageal pressure topography (EPT) offers a potential diagnostic advantage over conventional line tracings (CLT). However, high-resolution manometry (HRM) and EPT require increased equipment costs over conventional systems and evidence demonstrating a significant diagnostic advantage of EPT over CLT is limited. Our aim was to investigate whether the inter-rater agreement and/or accuracy of esophageal motility diagnosis differed between EPT and CLT. METHODS: Forty previously completed patient HRM studies were selected for analysis using a customized software program developed to perform blinded independent interpretation in either EPT or CLT (six pressure sensors) format. Six experienced gastroenterologists with a clinical focus in esophageal disease (attendings) and six gastroenterology trainees with minimal manometry experience (fellows) from three academic centers interpreted each of the 40 studies using both EPT and CLT formats. Rater diagnoses were assessed for inter-rater agreement and diagnostic accuracy, both for exact diagnosis and for correct identification of a major esophageal motility disorder. RESULTS: The total group agreement was moderate (=0.57; 95% CI: 0.56-0.59) for EPT and fair (=0.32; 0.30-0.33) for CLT. Inter-rater agreement between attendings was good (=0.68; 0.65-0.71) for EPT and moderate (=0.46; 0.43-0.50) for CLT. Inter-rater agreement between fellows was moderate (=0.48; 0.45-0.50) for EPT and poor to fair (=0.20; 0.17-0.24) for CLT. Among all raters, the odds of an incorrect exact esophageal motility diagnosis were 3.3 times higher with CLT assessment than with EPT (OR: 3.3; 95% CI: 2.4-4.5; P<0.0001), and the odds of incorrect identification of a major motility disorder were 3.4 times higher with CLT than with EPT (OR: 3.4; 2.4-5.0; P<0.0001). CONCLUSIONS: Superior inter-rater agreement and diagnostic accuracy of esophageal motility diagnoses were demonstrated with analysis using EPT over CLT among our selected raters. On the basis of these findings, EPT may be the preferred assessment modality of esophageal motility.

DOI: https://doi.org/10.1038/ajg.2015.159

Originally published at:
Carlson, Dustin A; Ravi, Karthik; Kahrilas, Peter J; Gyawali, C Prakash; Bredenoord, Arjan J; Castell, Donald O; Spechler, Stuart J; Halland, Magnus; Kanuri, Navya; Katzka, David A; Leggett, Cadman L;

Roman, Sabine; Saenz, Jose B; Sayuk, Gregory S; Wong, Alan C; Yadlapati, Rena; Ciolino, Jody D; Fox, Mark R; Pandolfino, John E (2015). Diagnosis of esophageal motility disorders: esophageal pressure topography vs. conventional line tracing. American Journal of Gastroenterology, 110(7):967-977.
DOI: https://doi.org/10.1038/ajg.2015.159

Diagnosis of esophageal motility disorders: esophageal pressure topography versus conventional line tracing.

Carlson, DA[1], Ravi, K[2], Kahrilas, PJ[1], Gyawali, CP[3], Bredenoord, AJ[4], Castell, DO[5], Spechler, SJ[6], Halland, M[2], Kanuri, N[3], Katzka, DA[2], Leggett, CL[2], Roman, S[7], Saenz, JB[3], Sayuk, GS[3], Wong, AC[1], Yadlapati, R[1], Ciolino, JD[8], Fox, MR[9], Pandolfino, JE[1]

[1] Division of Gastroenterology and Hepatology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; [2]Division of Gastroenterology, Mayo Clinic, Rochester, Minnesota, USA; [3]Division of Gastroenterology, Washington University School of Medicine, Saint Louis, Missouri, USA; [4]Department of Gastroenterology and Hepatology, Academic Medical Center Amsterdam, Amsterdam, The Netherland; [5]Department of Gastroenterology & Hepatology, Medical University of South Carolina, Charleston, South Carolina, USA; [6]Department of Medicine, VA North Texas Healthcare System, Dallas, Texas, USA; [7]Department of Digestive Physiology, Hospices Civils de Lyon, Lyon I University and Inserm 1032, Lyon, France; [8]Biostatistics Division, Division of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; [9]Division of Gastroenterology and Hepatology, University Hospital Zurich, Zurich, Switzerland

Corresponding Author:

John E Pandolfino

Northwestern University

Feinberg School of Medicine

Division of Gastroenterology and Hepatology

676 St Clair St, Suite 1400

Chicago, IL 60611-2951:

Email: j-pandolfino@northwestern.edu

Word Count: 3755

Key words: esophageal manometry, esophageal pressure topography, esophageal motility disorder

Abstract:

Background: Enhanced characterization of esophageal peristaltic and sphincter function provided by esophageal pressure topography (EPT) offers a potential diagnostic advantage over conventional line tracings (CLT). However, high-resolution manometry (HRM) and EPT require increased equipment costs over conventional systems and evidence demonstrating a significant diagnostic advantage of EPT over CLT is limited. Our aim was to investigate whether the inter-rater agreement and/or accuracy of esophageal motility diagnosis differed between EPT and CLT.

Methods: Forty previously-completed patient HRM studies were selected for analysis using a customized software program developed to perform blinded independent interpretation in either EPT or CLT (six pressure sensors) format. Six experienced gastroenterologists with a clinical focus in esophageal disease (attendings) and six gastroenterology trainees with minimal manometry experience (fellows) from three academic centers interpreted each of the 40 studies using both EPT and CLT formats. Rater diagnoses were assessed for inter-rater agreement and diagnostic accuracy, both for exact diagnosis and for correct identification of a major esophageal motility disorder.

Results: The total group agreement was moderate (κ = 0.57; 95% CI 0.56-0.59) for EPT and fair (κ = 0.32; 0.30-0.33) for CLT. Inter-rater agreement between attendings was good (κ = 0.68; 0.65-0.71) for EPT and moderate (κ = 0.46; 35 0.43-0.50) for CLT. Inter-rater agreement between fellows was moderate (κ = 0.48; 0.45-0.50) for EPT and poor to fair (κ = 0.20; 0.17-0.24) for CLT. Among all raters, the odds of an incorrect exact esophageal motility diagnosis were 3.3 times higher with CLT assessment than with EPT (OR 3.3; 95% CI 2.4-4.5; p<0.0001) and the odds of incorrect identification of a major motility disorder were 3.4 times higher with CLT than EPT (OR 3.4; 2.4-5.0; p<0.0001).

Conclusions: Superior inter-rater agreement and diagnostic accuracy of esophageal motility diagnoses was demonstrated with analysis using EPT over CLT among our selected raters. Based on these findings, EPT may be the preferred assessment modality of esophageal motility.

STUDY HIGHLIGHTS

WHAT IS CURRENT KNOWLEDGE

- Esophageal manometry is the primary method for evaluating for esophageal motility disorders and two formats are available: conventional line tracings and esophageal pressure topography.
- Though esophageal pressure topography and high resolution manometry systems potentially improve manometric interpretation, there is little evidence directly comparing the diagnostic effectiveness of esophageal pressure topography and conventional line tracings.

WHAT IS NEW HERE

- Among raters both with and without significant manometry interpretation experience, esophageal pressure topography provided superior inter-rater agreement and diagnostic accuracy of esophageal motility diagnoses over conventional line tracings.

Introduction

Esophageal manometry is the primary method for evaluating esophageal motor function and diagnosing esophageal motility disorders. These disorders were originally defined using a line tracing format that displayed pressure measurements acquired at short intervals (often three to five cm apart) along the length of the esophagus to define propagation speed and vigor of the peristaltic contraction. High resolution manometry (HRM), which utilizes pressure sensors spaced at one cm intervals throughout the entire length of the esophagus to generate esophageal pressure topography (EPT; Clouse plots), has been introduced as a possible advancement in evaluation of esophageal motility disorders based on a more refined spatiotemporal display of pressure measurements (1,2). In addition to offering a visual representation of esophageal motility, EPT provides objective metrics of esophageal function based on high resolution pressure data that form the basis of the EPT classification scheme of esophageal motility disorders, the Chicago Classification (3,4). However, HRM/EPT carries increased equipment and maintenance costs compared with conventional manometry assemblies and the direct evidence substantiating the diagnostic superiority of EPT analysis over the conventional line tracing (CLT) analysis is limited. Previous studies that compared the two display formats have focused on assessment of peristaltic integrity and/or success of bolus transit (findings without strong evidence to support alteration of clinical management) or included assessment of single swallows (not complete patient studies) by medical trainees (5-8). The reliability of esophageal motility assessment has been evaluated in each display method individually, but no head-to-head comparison has been made (9,10).

Though HRM/EPT is increasingly utilized in research and clinical practice, the use of conventional manometry with CLT assessment remains common. Illustrative of this, the European Achalasia trial, a large, multi-center, randomized controlled trial, utilized conventional manometry (with 6-10 sensors) to diagnose achalasia for inclusion into the study (11,12). Although this was likely related to limited access to EPT in some European countries, it does suggest that these investigators felt comfortable that CLT data were adequate for the diagnosis of achalasia despite existing data suggesting only moderate to good reproducibility among experts (9). We hypothesized that the agreement and accuracy for the diagnosis of esophageal motor disorders is greater for EPT than CLT. Our aim was to assess the inter-rater agreement of both techniques in a head-to-head comparison that optimized both approaches using a randomized, cross-over study design. Additionally, we also sought to determine which technique was most accurate by using technique-specific reference standards.

Methods

Subjects and study design

Experienced gastroenterologists with a clinical focus in esophageal disease (*attendings*) and gastroenterology trainees with minimal experience in esophageal manometric interpretation (*fellows*) from three academic medical centers (Northwestern University, Chicago, IL; Mayo Clinic, Rochester, MN; Washington University, St. Louis, MO) were invited to participate in the cross-over study. A total of seven attendings and seven fellows (two to three from each center) were initially invited to participate.

Raters were randomized to analysis order (i.e., EPT or CLT first). Randomization was blocked by institution and position (attending/fellow) in groups of two in a 1:1 allocation ratio. Raters provided informed consent and completed a baseline experience questionnaire assessing the number of manometries in each display format they had previously interpreted. Each rater watched an approximately 30-minute orientation recording discussing the study protocol, use of the analysis software, and classification criteria of esophageal motility disorders. Contents of the orientation, including classification criteria, were available to the raters throughout the study period. Each rater analyzed a set of the same 40 patient esophageal manometry studies displayed in both EPT and CLT formats separated in time by at least two weeks (see **Figure 1**). Rater analysis was completed over the study period spanning from 10/2013-11/2014. The protocol was approved by the Northwestern University Institutional Review Board.

Manometry studies and analysis

Manometry studies were previously performed between 2007 and 2013 as part of patients' standard evaluation. All manometries were done in a supine position after at least a 6 hour fast using a high-resolution catheter with a 4.2-mm outer-diameter, solid-state assembly with 36 circumferential sensors spaced at 1-cm intervals (Given Imaging, Duluth, Georgia, USA). The catheter was placed transnasally into the stomach and positioned to record from approximately the hypopharynx to the proximal stomach. The manometry study protocol included at least a 30-second baseline period and ten supine 5-ml liquid swallows.

Patient studies were selected by review of the manometry database at the Northwestern Esophageal center in reverse chronologic order to supply each of the Chicago Classification diagnoses (3). Studies were excluded if patients had previous esophageal surgery or intervention (e.g. dilation) or were technically limited. We intentionally included a preponderance of achalasia patients to assess the diagnostic effectiveness for the esophageal motility disorder with the most standardized treatment plan (13). Esophageal motility diagnoses were confirmed (i.e. the reference standard diagnosis) for each display method by assessment of each patient study and consensus agreement of two authors for the associated classification schemes: Chicago Classification/EPT reference standard diagnoses were confirmed by Arjan Bredenoord and John Pandolfino and the conventional classification scheme (with updates as noted)/CLT diagnoses were confirmed by Stuart Spechler and Don Castell (see **Tables 1A and 1B** (3,14-16). Patient clinical data, including symptomatology, endoscopic and imaging findings, and response to treatment was provided to the reference standard designators to settle discordant diagnoses if needed. The total numbers of included patient studies with each reference standard are listed in **Tables 2A** (EPT) and **2B** (CLT).

Studies were analyzed using Manoview version 3.0.1 (Given Imaging). Using manufacturer support, the provided analysis software was altered to allow viewing of each patient study with only the desired study display format (EPT *or* CLT, as illustrated in **Figure 3**). This ensured that raters could not toggle between display formats to utilize both techniques during analysis. For CLT, a six-sensor display was chosen to reflect the conventional classification criteria (14). Analysis software, de-identified manometry studies, and a copy of the orientation slides including the classification criteria were shared using a secure online drop-box. The sequence and file name of the manometry studies were randomly altered between each rater analysis. Pressure sensors for CLT analysis were placed prior to rater analysis by the study coordinator using the EPT plot to provide optimal sensor position for CLT analysis. A 6-cm eSleeve tracing, which simulates a Dent sleeve sensor, was also provided for CLT analysis. For EPT analysis, raters were required to individually place analysis landmarks (as instructed in the orientation session). Raters provided an esophageal motility diagnosis for each manometry study based on Chicago Classification for EPT and conventional criteria proposed for CLT as displayed in **Tables 1A** and **1B**, respectively (3,14-16).

Statistical analysis

Agreement between raters was calculated using Fleiss Kappa (κ) (17). Degree of agreement was interpreted based on kappa values as poor (0-0.2), fair (0.21-0.4), moderate (0.41-0.6), good (0.61-0.8), and excellent (0.81-1). Diagnostic accuracy was determined by the agreement with the reference standard for the respective classification scheme. Accuracy was determined both for *exact* agreement between rater and reference standard and for correct identification of any major esophageal motility disorder (e.g., rater diagnosis of distal esophageal spasm (DES) and reference standard for type III achalasia was considered an accurate identification of a major motility disorder; in contrast, rater diagnosis of DES and reference standard diagnosis of normal was considered inaccurate). Major motility disorders are labeled in **Tables 1A and 1B**. We felt the identification of any major motility disorder was an important distinction because it may indicate patients that are less likely labeled as functional dysphagia and more likely to undergo a change in management recommendation based on their manometry findings. Though misclassification *within* the major motility disorders can have clinical consequences, this distinction was reflected in the analysis of exact accuracy. To account for associations and dependency of outcomes within patients and within raters, between-group (EPT and CLT) comparisons were assessed using conditional logistic regression and generalized estimating equations were used to explore accuracy differences between attendings and fellows. Accuracy results are reported in terms of odds ratios (OR) for an incorrect diagnosis when assessed with CLT over EPT.

Results

Baseline rater characteristics

After randomization, one attending and one fellow withdrew from the study. A total of six attendings and six fellows completed the study protocol. Four attendings and three fellows analyzed EPT first and two attendings and three fellows analyzed CLT first. Baseline manometry experience characteristics are displayed in **Table 3**.

Inter-rater agreement

Agreement for esophageal motility diagnoses among all raters was moderate (κ = 0.57; 95% confidence interval 0.56 – 0.59) for EPT and fair (κ = 0.31; 0.30 – 0.33) for CLT. Agreement among attendings was good (κ = 0.68; 0.65 – 0.71) for EPT and moderate (κ = 0.46; 0.43 – 0.50) for CLT. Agreement among fellows was moderate (κ = 0.48; 0.45 – 0.50) for EPT and poor to fair (κ = 0.20; 0.17 – 0.24) for CLT. A pooled assessment of agreement and manner of disagreement are illustrated in **Tables 2A** and **2B**.

Agreement between all raters by individual esophageal motility diagnoses is displayed in **Table 4**. Between all raters with EPT, agreement was good to excellent for type I (κ = 0.82; 0.78 – 0.86) and type II (κ = 0.77; 0.73 – 0.81) achalasia and fair to moderate for type III achalasia (κ = 0.39; 0.35 – 0.43) and EGJ outflow obstruction (κ = 0.45; 0.41 – 0.49). With CLT, agreement among all raters was moderate to good (κ = 0.57; 0.54 – 0.62) for classic achalasia, but poor for atypical disorder of lower esophageal sphincter (LES) relaxation (κ = 0.10; 0.06 – 0.14). For patients with a normal motility diagnosis, agreement between raters was good (κ = 0.53; 0.49 – 0.57) with EPT and fair (κ = 0.25; 0.23 – 0.29) with CLT.

When inter-rater agreement was assessed by motility diagnosis and by positions (i.e., attending/fellows), agreement between attendings with EPT was good to excellent for type I (κ = 0.82; 0.74 – 0.90) and type II (κ = 0.80; 0.72 – 0.88) achalasia and moderate to good for normal (κ = 0.58; 0.50 – 0.66). Agreement between attendings with CLT was good to excellent for classic achalasia (κ = 0.77; 0.69 – 085) and moderate for normal (κ = 0.37; 0.30 – 0.46). Agreement for all diagnoses by position is reported in **Tables 5A (attendings) and 5B (fellows)**.

Diagnostic accuracy

When comparing rater diagnoses with the reference standard diagnoses for exact agreement of esophageal motility diagnoses, among all raters, the odds of an incorrect diagnosis for CLT was 3.3 times the odds of an incorrect diagnosis with EPT interpretation (OR 3.3, 95%CI 2.4-4.5, $p<0.0001$). The odds of an incorrect exact diagnosis was

3.9 times higher with CLT than EPT with attending raters (OR 3.9, 2.3 - 6.7, p<0.0001) and 6.3 times higher with CLT than EPT with fellow raters (OR 6.3, 3.7 – 10.9, p<0.0001). Odds of an incorrect exact diagnosis were 1.9 times higher for fellows than for attendings (OR 1.9, 1.5-2.4, p<0.0001), however the interaction between position and assessment method was not significant (p=0.14).

When loosening the definition of diagnostic accuracy and assessing for correct identification of a *major* esophageal motility disorder, the odds of incorrect diagnosis of a major motility disorder among all raters were 3.4 times higher with CLT interpretation than with EPT (OR 3.4; 2.4-5.0; p<0.0001). The odds of an incorrect diagnosis of a major motility disorder was 5.1 times higher with CLT than EPT with attending raters (OR 5.1, 2.3 – 11.4, p<0.0001) and 5.1 times higher with CLT than EPT with fellow raters (OR 5.1, 2.3 – 11.1, p<0.0001). The odds of an incorrect major motility disorder diagnosis did not statistically differ (at the 5% level) between attendings and fellows (p=0.09).

Pooled accuracy results for all raters are illustrated in **Tables 2A** and **2B** and in **Figure 3**.

Discussion

The aim of this study was to compare the inter-rater agreement and accuracy of the two primary manometric display formats, EPT and CLT, for the diagnosis of esophageal motility disorders. Reliability and validity are hallmarks of any diagnostic test, thus we assessed the agreement between multiple raters, some with extensive and some with limited experience in manometry interpretation, and diagnostic accuracy, as determined by comparison with a reference standard diagnostic classification agreed upon by expert physicians that published the esophageal motility classification schemes. We demonstrated a higher level of agreement among multiple raters and significantly better diagnostic accuracy for esophageal motility diagnoses when raters used EPT as opposed to CLT.

Beyond diagnostic accuracy, another important feature of a diagnostic test is to provide information that ultimately affects clinical management decisions. The clinical significance of some esophageal motility *classifications* (such as hypertensive peristalsis in the Chicago Classification or isolated hypertensive LES in the conventional classification) remains unclear (3,14). In acknowledgment of this, the Chicago Classification has recently been updated and has removed hypertensive peristalsis and rapid contraction with normal latency from its classification scheme (4). Additionally, the clinical significance of *misclassification* of diagnosis (e.g. distinguishing normal peristalsis from weak peristalsis or IEM) may be marginal if it does not affect patient management decisions. Though assessment of inter-rater reliability for certain diagnoses was somewhat limited by small sample numbers, we intentionally included an excess of achalasia patients. While this may somewhat limit our study's generalizability to a community practice, we thought it was imperative to assess the diagnosis of this important esophageal motility disorder with a generally standardized treatment plan (13). Overall inter-rater agreement was moderate to good among classic achalasia diagnoses with CLT, but good to excellent for type I and type II achalasia with EPT. Among only attendings, inter-rater agreement was similar for the diagnosis of classic achalasia with CLT and diagnosis of type I and type II achalasia with EPT. Reliable diagnosis of achalasia is paramount because the recommended treatment plan for achalasia often involves a recommendation for specific surgical management (Heller myotomy) or pneumatic dilation (10). Thus, among physicians experienced in manometry interpretation, classic achalasia may be reliably diagnosed with either EPT or CLT. However, evaluation of other disorders of impaired LES relaxation, such as spastic achalasia (type III achalasia and/or atypical disorder of LES relaxation), may be limited with CLT as attending raters exhibited good agreement for type III achalasia and EGJ-outflow obstruction with EPT, but poor for atypical disorders of LES relaxation with CLT. Additionally, we assessed the ability for each manometric methodology to identify patients with a *major* esophageal motility disorder; i.e. a designation that would likely impact on treatment recommendations or direct further clinical evaluation based on their manometric diagnosis. In doing so, we found that EPT provided significantly improved diagnostic accuracy over CLT for this distinction with the odds of an incorrect diagnosis being greater than threefold higher using CLT as opposed to EPT.

Previous studies have reported fair to excellent inter-rater reliability and/or agreement of esophageal motility disorders with both EPT and CLT assessment, though no head-to-head comparison of reliability has been completed between the two methods. Nayar et al. evaluated CLT diagnoses of 72 patients between eight raters (three highly experienced, three moderately experienced, and two inexperienced raters, defined by > 1000, 100-999, and < 100 previous manometry interpretations, respectively) (9). Similar to the results of our study, the authors observed moderate agreement ($\kappa = 0.44$) between highly experienced raters, and fair agreement ($\kappa = 0.3$) between inexperienced raters for diagnoses based on CLT. The agreement between all and highly experienced raters improved to good ($\kappa = 0.68$ for both) when only evaluating for agreement between achalasia and normal. Recently, an evaluation of inter-observer agreement for EPT interpretation of 40 patient studies involving 36 raters (nine with > 400 and 17 with < 200 previous HRM analyses) demonstrated an overall moderate agreement between all raters ($\kappa = 0.41$) and highly experienced raters ($\kappa = 0.51$) (10). However, the provided analysis program for that online study did not provide full Chicago Classification metrics such as the IRP or distal contractile integral (DCI). Another study evaluating EPT interpretation between five raters analyzing 30 patients demonstrated good to very good (Krippendorff's $\alpha > 0.75$) agreement on distinction between non-achalasia and achalasia patients and between achalasia subtypes (18). Other studies assessing agreement between EPT metrics have demonstrated good to excellent agreement between small numbers of raters (N = 2-4) (19,20).

Few studies have compared the diagnostic accuracy of esophageal motility disorders between EPT and CLT. An early study by Clause and Staiano using 21-channel water-perfused manometry compared the accuracy of esophageal motility diagnoses using topographic representation versus limited, 4-channel, line tracings (sensors at the LES and 3, 8, and 13 cm above the LES) for 212 consecutive patients referred for manometry (21). Using clinical information to settle diagnostic discrepancies between display formats, diagnostic interpretation of 10 water swallow-studies between the two display formats was compared. They concluded that topographic plots allowed better localization and assessment of the LES and improved diagnosis of achalasia over line tracing assessment with limited pressure sensors. However, since this study, numerous advances have been made including development of EPT-specific software and metrics, as well as updates in classification schemes.

Other studies have reported improved accuracy of esophageal motility interpretation with EPT over line tracings in evaluation of single swallows performed by medical trainees (fellows, residents, and mostly medical students) (7,8). However, evaluation of only medical trainees and single swallow limits generalizability to clinical practice. To simulate an actual clinical manometry assessment, we utilized commercially available analysis software (albeit altered to control for study conditions) and included raters with a range of experience in manometric interpretation. It is possible that our inclusion of significantly experienced, specialist physicians could limit the generalizability to community practice; however, the inclusion of inexperienced fellows as raters suggests the benefits of EPT over CLT across the full spectrum of experience in manometry interpretation.

A limitation to our study include was that each patient study was performed with HRM catheters, as is the practice at our institution, and only the analysis software was altered to allow analysis with the desired output format. While the results could be suggestive of an indirect comparison between HRM and conventional manometry assemblies, our objective was to compare assessment with EPT and CLT, not necessarily high-resolution and conventional manometry. Furthermore, we did not want to subject patients to two separate transnasal manometry catheter intubations. Though line tracing catheters and displays are available with more than six pressure sensors, we chose to use the six sensors display consistent with recommendations for the acquisition and analysis of conventional manometry (14). Patient studies were not completed specifically for this study, but were selected retrospectively for inclusion based on Chicago Classification diagnoses, which may introduce a bias. However, our study design may have actually enhanced CLT interpretation because EPT-defined landmarks were used to place the CLT pressure sensors; this optimally positioning the LES sensor within the EGJ high-pressure zone during the deglutitive period (as opposed to utilizing a pull-through EGJ localization method, which may be susceptible to misidentifying EGJ pressure during swallow-induced esophageal shortening). Additionally, an electronic-sleeve sensor was provided to enhance CLT assessment of EGJ pressure and relaxation. Overall, we believe that this study design minimized the potential for technical limitations in the CLT studies and focused the analysis on a fair comparison of the two analysis formats.

Another limitation was the reliance on the reference standard diagnoses for determining diagnostic accuracy. A true *gold standard* for esophageal motility diagnosis is difficult to establish, and as also observed in previous studies, there is variability among even experienced raters in determining esophageal motility diagnoses (9,10). We attempted to create a best possible reference standard based on the manometric output using the consensus assessment of two authors of the primary classification schemes for each display format; though our focus was primary on manometry assessment, reference standard determination was aided by clinical information when disagreement occurred. It may be noted that diagnostic agreement *between* the two formats for seemingly corresponding diagnoses was not consistent. While extrapolating esophageal motility diagnoses from one display/classification scheme to another may have clinical relevance, it was not the aim of this study and a secondary analysis will address this observation.

There are several explanations to account for our findings of improved inter-rater agreement and diagnostic accuracy with EPT over CLT. The pictorial representation provided with EPT may be more easily interpreted through pattern recognition than the less intuitive display with line tracings. Additionally, the automated, quantitative metrics that are provided by EPT (e.g. IRP and DCI) lend themselves to algorithm-based diagnosis. Assessment of deglutitive LES relaxation is imperative for esophageal motility diagnosis, the availability of the IRP for the assessment of deglutitive LES relaxation with EPT analysis likely provided an advantage over the measurement of residual LES pressures with CLT. We did, however, require raters to place interpretation landmarks during their analysis for generation of the EPT automated landmarks. This cannot be done with CLT as the information is based on the six sensors and these were optimally positioned using EPT landmarks. As we did not ask raters to record measurements or provide classifications of individual swallow, the mechanism of disagreement and/or inaccuracy is difficult to determine. Additionally, there are other unmeasured rater-specific biases that could have influenced the results. We did not track time and/or individual effort spent on each interpretation modality, nor did we identify the timeframe of previous manometry interpretations (i.e. when experience was obtained), and thus may have reflected raters' more recent practice (particularly as most have shifted to EPT analysis in their clinical practice). Finally, though we attempted to include fellows with minimal previous manometry interpretation experience, we may not have accounted for their instruction during training (outside of independent manometry interpretation), which may have favored EPT over CLT analysis. However, the superior interpretation among fellows with EPT over CLT may support the ease of learning manometry based on pattern recognition afforded with EPT (8).In conclusion, we incorporated a cross-over study design of the two manometric display formats utilizing customized analysis software and demonstrated that EPT provided more reliable and more accurate interpretation of esophageal motility diagnoses than CLT; these findings were consistent for both specialist attending physicians from tertiary referral centers with experience in interpreting manometric studies and for fellows with little or no manometry interpretation experience. While our rater and patient selection may limit the generalizability to community practice, within the limits of our study, our findings suggest that interpretation of manometry studies using EPT format may be considered the preferred method for evaluating esophageal motility. Thus, use of EPT should be considered for recruitment of homogeneous patient groups in clinical trials and designation of esophageal motility diagnoses in clinical practice, especially when directing utilization of invasive diagnostic or therapeutic procedures.

Acknowledgements: We would like to thank Dave Costarella (Given Imaging) for technical support producing the specialized analysis software.

Conflict of interest

Guarantor of the article: John E. Pandolfino, MD

Specific author contributions: Dustin A. Carlson and John E. Pandolfino developed the study concept and design, acquired and analyzed the data, supervised the study, and drafted the manuscript. JEP also provided reference standard diagnoses. Karthik Ravi and C. Prakash Gyawali acted as study center coordinators and participated as raters. Peter J. Kahrilas participated as a rater and helped finalize the manuscript. Donald O. Castell and Stuart J. Spechler provided reference standard diagnoses, reviewed the study orientation, and reviewed the manuscript. Arjan J. Bredenoord provided reference standard diagnoses and approved the final manuscript. Jody D. Ciolino analyzed the data and assisted finalizing the manuscript. Mark R. Fox developed the study concept and assisted finalizing the manuscript. Magnus Halland, Navya Kanuri, David A Katzka, Cadman L. Leggett, Sabine Roman, Jose Saenz, Gregory S. Sayuk, Alan Wong, and Rena Yadlapati participated as raters and approved the final manuscript.

Financial support*:* This work was supported by T32 DK101363 (JEP) and RO1 DK079992 (JEP) from the Public Health service.

Potential competing interests*:* Arjan J. Bredenoord [Given Imaging/Covidien (research)]. Mark R. Fox [Given Imaging/Covidien (consulting, education)]. C. Prakash Gyawali [Given Imaging/Covidien (consulting, research)]. John E. Pandolfino [Given Imaging/Covidien (consulting, education)]. Sabine Roman [Given Imaging/Covidien (consulting, education, research, advisory board)]. There are no other conflicts for remaining authors: Dustin A. Carlson, Donald O. Castell, Jody D. Ciolino, Magnus Halland, Peter J. Kahrilas, Navya Kanuri, David A. Katzka, Cadman L. Leggett, Karthik Ravi, Jose B. Saenz, Gregory S. Sayuk, Stuart J. Spechler, Alan C. Wong, and Rena Yadlapati.

References

1. Clouse, R.E. & Staiano, A. Topography of the esophageal peristaltic pressure wave. *Am J Physiol* 1991; 261: G677-684.
2. Fox, M.R. & Bredenoord, A.J. Oesophageal high-resolution manometry: moving from research into clinical practice. *Gut* 2008; 57: 405-423.
3. Bredenoord, A.J., Fox, M., Kahrilas, P.J.*, et al.* Chicago Classification Criteria of Esophageal Motility Disorders Defined in High Resolution Esophageal Pressure Topography. *Neurogastroenterol Motil* 2012.
4. Kahrilas, P.J., Bredenoord, A.J., Fox, M.*, et al.* The Chicago Classification of esophageal motility disorders, v3.0. *Neurogastroenterol Motil* 2014.
5. Fox, M., Hebbard, G., Janiak, P.*, et al.* High-resolution manometry predicts the success of oesophageal bolus transport and identifies clinically important abnormalities not detected by conventional manometry. *Neurogastroenterol Motil* 2004; 16: 533-542.
6. Xiao, Y., Kahrilas, P.J., Kwasny, M.J.*, et al.* High-Resolution Manometry Correlates of Ineffective Esophageal Motility. *Am J Gastroenterol* 2012.
7. Grubel, C., Hiscock, R. & Hebbard, G. Value of spatiotemporal representation of manometric data. *Clin Gastroenterol Hepatol* 2008; 6: 525-530.
8. Soudagar, A.S., Sayuk, G.S. & Gyawali, C.P. Learners favour high resolution oesophageal manometry with better diagnostic accuracy over conventional line tracings. *Gut* 2012; 61: 798-803.
9. Nayar, D.S., Khandwala, F., Achkar, E.*, et al.* Esophageal manometry: assessment of interpreter consistency. *Clin Gastroenterol Hepatol* 2005; 3: 218-224.
10. Fox, M.R., Pandolfino, J.E., Sweis, R.*, et al.* Inter-observer agreement for diagnostic classification of esophageal motility disorders defined in high-resolution manometry. *Dis Esophagus* 2014.
11. Boeckxstaens, G.E., Annese, V., des Varannes, S.B.*, et al.* Pneumatic dilation versus laparoscopic Heller's myotomy for idiopathic achalasia. *N Engl J Med* 2011; 364: 1807-1816.
12. Rohof, W.O., Salvador, R., Annese, V.*, et al.* Outcomes of treatment for achalasia depend on manometric subtype. *Gastroenterology* 2013; 144: 718-725; quiz e713-714.
13. Vaezi, M.F., Pandolfino, J.E. & Vela, M.F. ACG clinical guideline: diagnosis and management of achalasia. *Am J Gastroenterol* 2013; 108: 1238-1249; quiz 1250.
14. Spechler, S.J. & Castell, D.O. Classification of oesophageal motility abnormalities. *Gut* 2001; 49: 145-151.
15. Agrawal, A., Hila, A., Tutuian, R.*, et al.* Clinical relevance of the nutcracker esophagus: suggested revision of criteria for diagnosis. *J Clin Gastroenterol* 2006; 40: 504-509.
16. Blonski, W., Vela, M., Safder, A.*, et al.* Revised criterion for diagnosis of ineffective esophageal motility is associated with more frequent dysphagia and greater bolus transit abnormalities. *Am J Gastroenterol* 2008; 103: 699-704.
17. Fleiss, J.L. Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin* 1971; 76: 378-382.
18. Hernandez, J.C., Ratuapli, S.K., Burdick, G.E.*, et al.* Interrater and intrarater agreement of the chicago classification of achalasia subtypes using high-resolution esophageal manometry. *Am J Gastroenterol* 2012; 107: 207-214.
19. Sweis, R., Anggiansah, A., Wong, T.*, et al.* Normative values and inter-observer agreement for liquid and solid bolus swallows in upright and supine positions as assessed by esophageal high-resolution manometry. *Neurogastroenterol Motil* 2011; 23: 509-e198.
20. Singh, E.R., Rife, C., Clayton, S.*, et al.* Interobserver Variability in Esophageal Body Measurements With High-resolution Manometry Among New Physician Users. *J Clin Gastroenterol* 2013; 47: e12-16.
21. Clouse, R.E., Staiano, A., Alrakawi, A.*, et al.* Application of topographical methods to clinical esophageal manometry. *Am J Gastroenterol* 2000; 95: 2720-2730.

1A: Chicago Classification	
Motility diagnosis	**Diagnostic Criteria**
Type I achalasia (Classic Achalasia)*	• Mean IRP > 15mmHg AND • 100% failed peristalsis
Type II achalasia (Achalasia with panesophageal pressurization)*	• Mean IRP > 15mmHg AND • Panesophageal pressurization with ≥20% of swallows AND • No evidence of normal peristalsis
Type III achalasia (Spastic achalasia)*	• Mean IRP ≥ 15mmHg AND • No normal peristalsis AND • Preserved fragments of distal contraction OR ≥ 20% of swallows with premature (DL < 4.5s) contractions
EGJ outflow obstruction*	• Mean IRP ≥ 15mmHg • Not meeting other criteria for any achalasia subtype
All of the following diagnoses have normal deglutitive LES relaxation (IRP < 15 mmHg)	
Distal esophageal spasm*	• ≥ 20% of swallows with premature contractions (DL ≤4.5s)
Jackhammer esophagus*	• At least one swallow DCI > 8000 mmHg•s•cm
Absent peristalsis*	• 100% of swallows with failed peristalsis
Weak peristalsis	• >20% swallows with large (>5cm axial length) breaks in the 20 mmHg isobaric contour OR • >30% swallows with small (2-5cm axial length) breaks in the 20 mmHg isobaric contour
Rapid contractions with normal latency	• CFV > 9s/cm with >20% of swallows AND DL >4.5s
Hypertensive peristalsis	• Mean DCI > 5000 mmHg•s•cm, but not meeting criteria for Jackhammer esophagus
Normal	• Not achieving any of the above diagnostic criteria

1B: Conventional Criteria	
Motility diagnosis	**Diagnostic Criteria**
Classic achalasia*	• Incomplete LES relaxation w/ swallow (residual LES pressure > 8 mmHg) AND • Aperistalsis
Atypical disorders of LES relaxation*	• Incomplete LES relaxation w/ swallow • Some preserved esophageal peristalsis • Complete LES relaxation of *inadequate duration*
All of the following diagnoses have complete deglutitive LES relaxation (to a level <8 mmHg above gastric pressure)	
Diffuse esophageal spasm*	• May have low, normal, or high LES pressure • Simultaneous contractions in >10% of swallows AND mean simultaneous contraction amplitude > 30mmHg
Nutcracker esophagus w/ hypertensive LES*†	• Mean resting LES pressure >45mmHg in midrespiration AND • Peristaltic wave amplitude >220mmHg
Nutcracker esophagus*†	• Mean distal esophageal peristaltic wave amplitude > 220 mmHg
Isolated hypertensive LES	• Mean resting LES pressure >45mmHg in midrespiration

Ineffective esophageal motility†	• >50% of swallows w/ any of the following:
	    • Distal esophageal wave amplitude <30 mmHg
	    • Simultaneous contractions w/ waves < 30 mmHg
	    • Failed peristalsis
	• Absent peristalsis

Normal	• Basal LES pressure: 10-45 mmHg at mid-respiratory pressure • Distal wave amplitude: 30-180 mmHg • Wave progression: Peristalsis progressing from UES to LES at 2-8 cm/s

Table 1. Classification criteria utilized for esophageal pressure topography (2A: Chicago Classification) and conventional line tracing (2B: Conventional criteria) (3, 14-16). *****Indicates *major* esophageal motility disorders. †Indicates criteria updated from the original proposed classification scheme (14-16). IRP – integrated relaxation pressure. DL – distal latency. EGJ – esophagogastric junction. LES – lower esophageal sphincter. DCI – distal contractile integral. CFV – contractile front velocity. UES – upper esophageal sphincter.

EPT versus CLT

2A. EPT		**Reference Standard**											
		Type I Achalasia (4)	Type II Achalasia (6)	Type III Achalasia (4)	EGJOO (4)	DES (3)	Jackhammer (3)	Absent peristalsis (2)	Weak Peristalsis (5)	Rapid (2)	HTN Peristalsis (2)	Normal (5)	Total (40)
Raters	Type I Achalasia	41	2	1									
	Type II Achalasia	7	69	11		1		1					
	Type III Achalasia			29	6	4			1				
	EGJOO			3	31	1	1		4	2		4	
	DES				1	17			1	3		1	
	Jackhammer			2			30			7	2		
	Absent Peristalsis						1	23	1				
	Weak Peristalsis				1	5			43			4	
	Rapid				1	1			2	7		5	
	HTN Peristalsis		1	1		3	4			4	19	2	
	Normal			1	8	4			8	1	3	44	
Total		48	72	48	48	36	36	24	60	24	24	60	480

2B. CLT		**Reference Standard**								
		Classic Achalasia (11)	ADLESR (17)	DES (2)	Nutcracker (2)	Nutcracker + HTN LES (0)	Isolated HTN LES (1)	IEM (3)	Normal (4)	Total (40)
Raters	Classic Achalasia	91	14	1						
	ADLESR	15	57		3		3	1	9	
	DES	1	35	17	1		2		7	
	Nutcracker		6	2	11				1	
	Nutcracker + HTN LES	1	10	1	3		1		1	
	Isolated HTN LES	1	19	1	1		2	2	7	
	IEM	23	19					27	1	
	Normal		44	2	5		4	6	22	
Total		132	204	24	24	0	12	36	48	480

Table 2. Contingency tables of rater and reference standard esophageal motility diagnoses in esophageal pressure topography (EPT, 2A) and conventional line tracing (CLT, 2B). The number of studies designated by the reference standards for each format are included in the second rows of each table (N). The values representing the pooled number of patients designated by the raters into each diagnoses (i.e. each patient study obtained a total of 12 rater diagnoses) and does not account for repeated measures within individual patients. The manner of agreement and disagreement *between raters* within esophageal motility diagnoses can be assessed *across rows*. *Accuracy (*shaded boxes)*,* and the manner of inaccuracy, compared with the reference standard can be assessed *down columns*. EGJOO – esophagogastric junction outflow obstruction. DES – distal esophageal spasm. Rapid – Rapid contractions with normal latency. HTN – hypertensive. LES – lower esophageal sphincter. ADLESR – atypical disorder of LES relaxation. IEM – ineffective esophageal motility.

	Attending (N=6)		**Fellow** (N=6)	
Previous manometries interpreted	**EPT**	**CLT**	**EPT**	**CLT**
<25			6	6
25-100		1		
101-499	1	1		
>500	5	4		

Table 3. Rater experience. Values represent the number of raters at each experience level. Experience level was defined by the number of manometries interpreted in each display format prior to the study onset. EPT – esophageal pressure topography. CLT – conventional line tracings.

Esophageal pressure topography			Conventional line tracings	
Motility diagnosis	**K (95% CI)**		**Motility diagnosis**	**K (95% CI)**
Type I achalasia	0.82 (0.78 - 0.86)		Classic achalasia	0.58 (0.54 - 0.62)
Type II achalasia	0.77 (0.73 - 0.81)			
Type III achalasia	0.39 (0.35 - 0.43)		ADLESR	0.10 (0.06 - 0.14)
EGJOO	0.45 (0.41 - 0.49)			
DES	0.32 (0.28 - 0.35)		DES	0.28 (0.25 - 0.32)
Jackhammer	0.62 (0.58 - 0.66)		Nutcracker	0.23 (0.19 - 0.27)
Absent peristalsis	0.87 (0.37 – 0.45)		Nutcracker w/ HTN LES	0.15 (0.11 - 0.19)
HTN peristalsis	0.41 (0.37 - 0.45)			
Rapid contraction w/ normal latency	0.18 (0.14 - 0.21)		Isolated HTN LES	0.09 (0.05 - 0.12)
Weak peristalsis	0.55 (0.51 - 0.59)		IEM	0.50 (0.47 - 0.54)
Normal	0.53 (0.49 - 0.57)		Normal	0.25 (0.21 - 029)

Table 4. Inter-rater agreement between all raters by esophageal motility diagnoses. EGJOO – esophagogastric junction outflow obstruction. DES – distal esophageal spasm. HTN – hypertensive. LES – lower esophageal sphincter. ADLESR – Atypical disorder of LES relaxation. IEM – ineffective esophageal motility.

5A: Attendings				
Esophageal pressure topography			**Conventional line tracings**	
Motility diagnosis	**K (95% CI)**		**Motility diagnosis**	**K (95% CI)**
Type I achalasia	0.82 (0.74 - 0.90)		Classic achalasia	0.77 (0.69 - 0.85)
Type II achalasia	0.80 (0.72 - 0.88)			
Type III achalasia	0.65 (0.57 - 0.73)		ATDLESR	0.12 (0.04 - 0.20)
EGJOO	0.61 (0.53 - 0.69)			
DES	0.61 (0.53 - 0.69)		DES	0.38 (0.30 - 0.46)
Jackhammer	0.68 (0.60 - 0.76)		Nutcracker	0.33 (0.25 - 0.41)
Absent peristalsis	0.92 (0.84 – 1.0)		Nutcracker w/ HTN LES	0.29 (0.21 - 0.37)
HTN peristalsis	0.59 (0.51 - 0.67)			
Rapid contraction w/ normal latency	0.24 (0.16 - 0.32)		Isolated HTN LES	0.07 (-0.01 - 0.15)
Weak peristalsis	0.79 (0.71 - 0.87)		IEM	0.69 (0.61 - 0.77)
Normal	0.58 (0.50 - 0.66)		Normal	0.38 (0.30 - 0.46)

5B: Fellows				
Esophageal pressure topography			**Conventional line tracings**	
Motility diagnosis	**K (95% CI)**		**Motility diagnosis**	**K (95% CI)**
Type I achalasia	0.80 (0.72 - 0.88)		Classic achalasia	0.38 (0.30 - 0.46)
Type II achalasia	0.72 (0.64 - 0.80)			
Type III achalasia	0.17 (0.09 - 0.25)		ATDLESR	0.01 (-0.07 - 0.09)
EGJOO	0.29 (0.21 - 0.37)			
DES	0.05 (-0.03 - 0.13)		DES	0.21 (0.13 - 0.29)
Jackhammer	0.57 (0.49 - 0.65)		Nutcracker	0.03 (-0.05 - 0.11)
Absent peristalsis	0.82 (0.74 – 0.90)		Nutcracker w/ HTN LES	0.11 (0.03 - 0.19)
HTN peristalsis	0.26 (0.18 - 0.34)			
Rapid contraction w/ normal latency	0.12 (0.04 - 0.20)		Isolated HTN LES	0.06 (-0.02 - 0.14)
Weak peristalsis	0.40 (0.32 - 0.48)		IEM	0.42 (0.34 - 0.50)
Normal	0.45 (0.37 - 0.53)		Normal	0.23 (0.15 - 0.31)

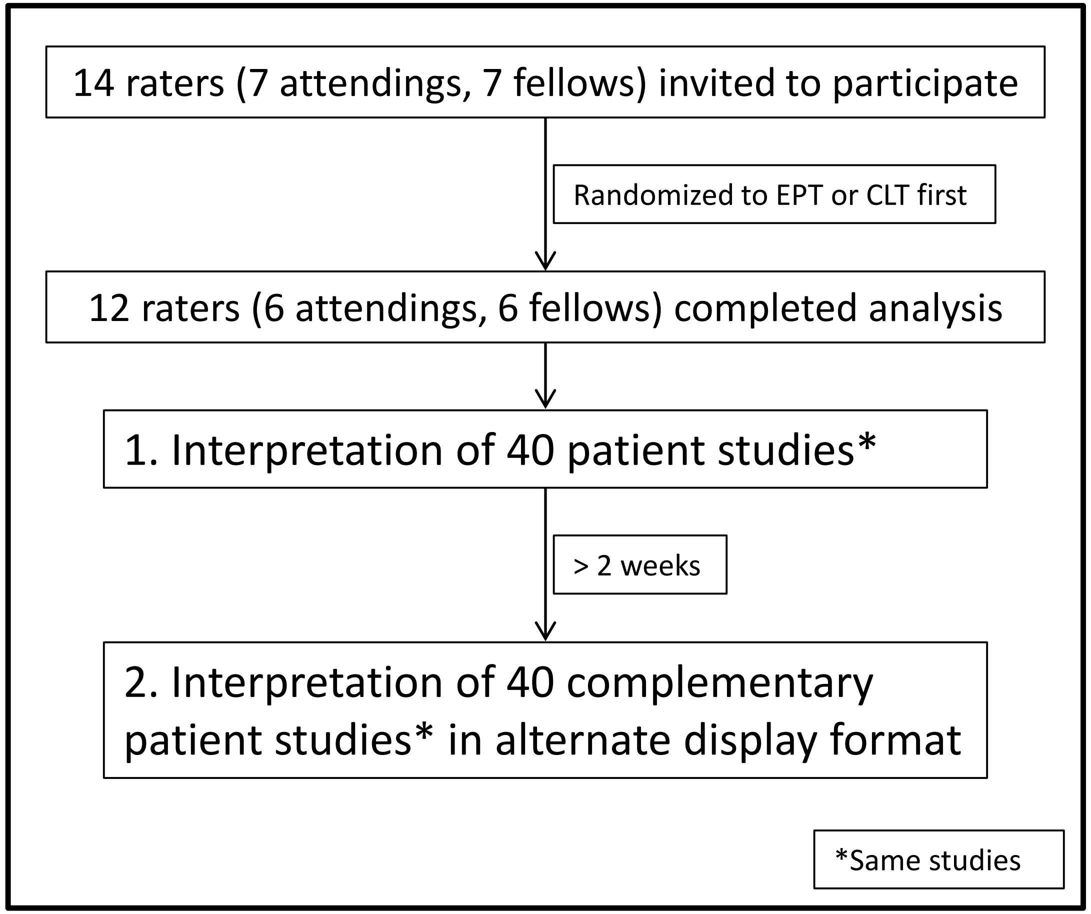
Table 5. Inter-rater agreement between attendings (A) and fellows (B) by esophageal motility diagnoses. EGJOO – esophagogastric junction outflow obstruction. DES – distal esophageal spasm. HTN – hypertensive. LES – lower esophageal sphincter. ADLESR – Atypical disorder of LES relaxation. IEM – ineffective esophageal motility.

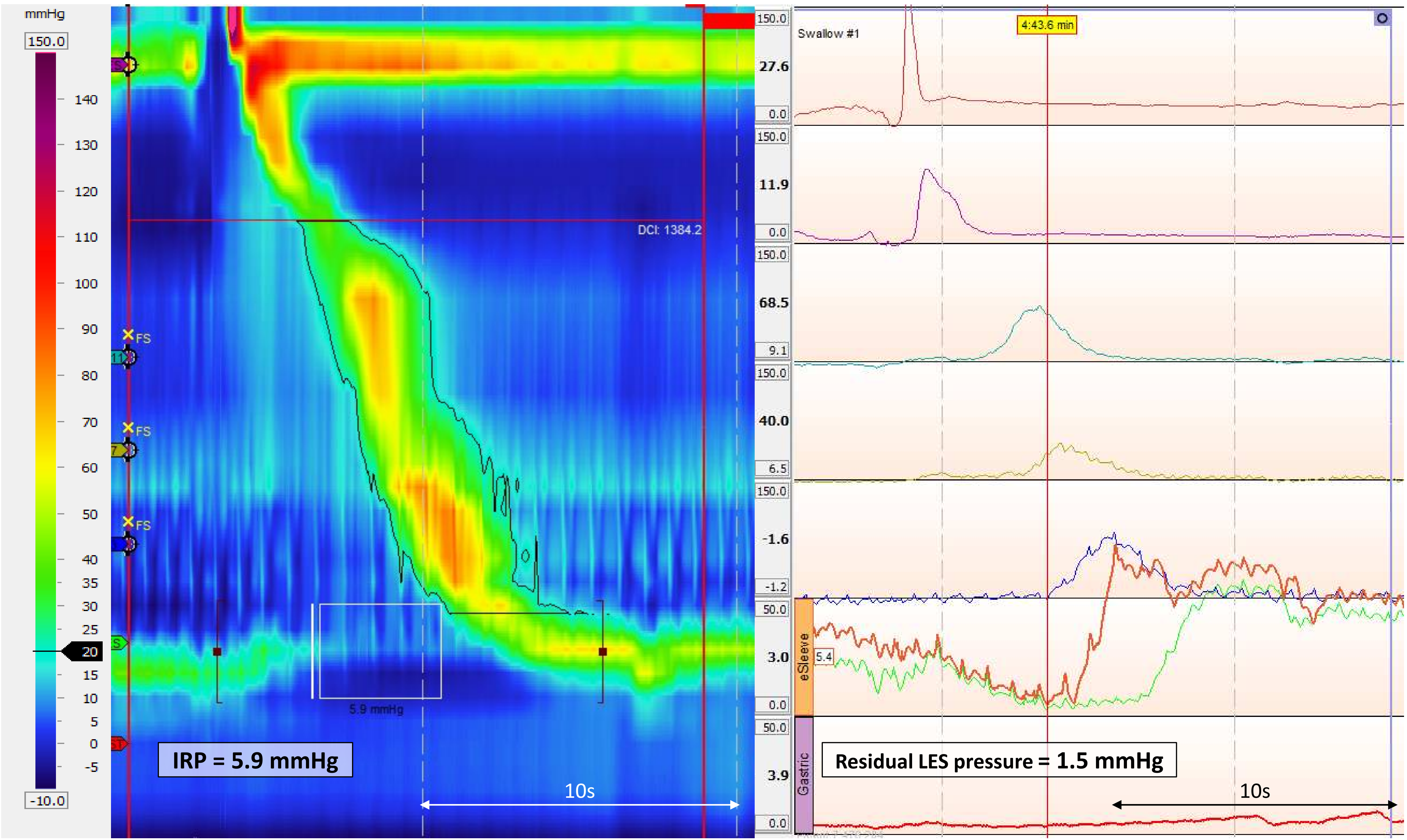
Figure 1. Study schema. Raters were randomized to the order of analysis (i.e., either esophageal pressure topography *or* conventional line tracing first). At least two weeks after completion of the first analysis, each rater analyzed the same 40 patient studies (re-ordered and re-coded) in the alternate display format than his/her first analysis.

Figure 2. Examples of manometry analysis software. Examples of single swallows from two patients included in the study are demonstrated in both esophageal pressure topography (EPT, left) and conventional line tracing (CLT, right). For CLT analysis, studies included pressure sensors placed in the stomach, through the center of the lower esophageal sphincter (LES) in the deglutitive window, at 3, 8, and 13 cm above the EGJ, and at the upper esophageal sphincter. Deglutitive LES relaxation was measured in EPT with the 4-second integrated relaxation pressure (IRP, white boxes) and in CLT by the residual LES pressure (difference between the LES nadir pressure and the gastric pressure). Both the single sensor LES pressure (green line) and 6cm eSleeve pressure (orange line) were provided for LES assessment in CLT. **2A) Normal swallow:** This patient was diagnosed as normal by both EPT and CLT reference standards, diagnosed as normal by 12/12 raters with EPT, and 9/12 by CLT [one fellow diagnosed the patient as ineffective esophageal motility (IEM) and two fellows as isolated hypertensive LES]. **2B) Achalasia**: This patient was diagnosed as type I achalasia (EPT) and classic achalasia (CLT) by the reference standards, type I achalasia by 11/12 raters with EPT (one fellow diagnosed as type II achalasia), and classic achalasia by 4/12 raters with CLT (4 attendings and four fellows diagnosed the patient as IEM).

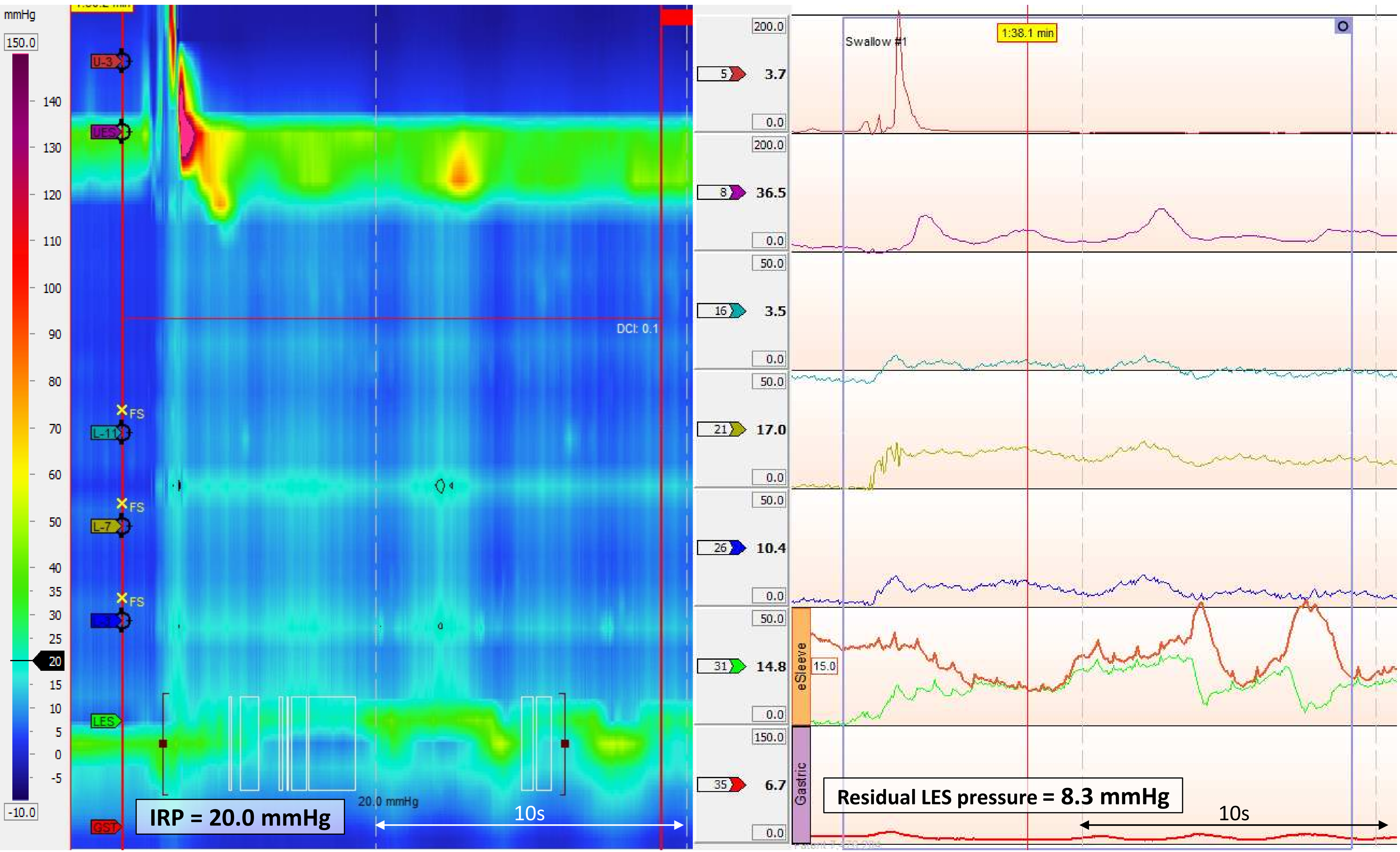
Figure 3. Diagnostic accuracy of esophageal motility disorders. Pooled assessment of all rater patient diagnoses by comparison with the reference standard diagnosis. Numbers represent the number of patients in each category. Disagreement with the reference standard on the exact diagnosis, but correct identification of a major motility disorder was considered **mild inaccuracy**. Incorrect identification of a major motility disorder was considered a **major inaccuracy**.

Figure 1


14 raters (7 attendings, 7 fellows) invited to participate
Randomized to EPT or CLT first
12 raters (6 attendings, 6 fellows) completed analysis
1. Interpretation of 40 patient studies*
> 2 weeks
2. Interpretation of 40 complementary patient studies* in alternate display format
*Same studies

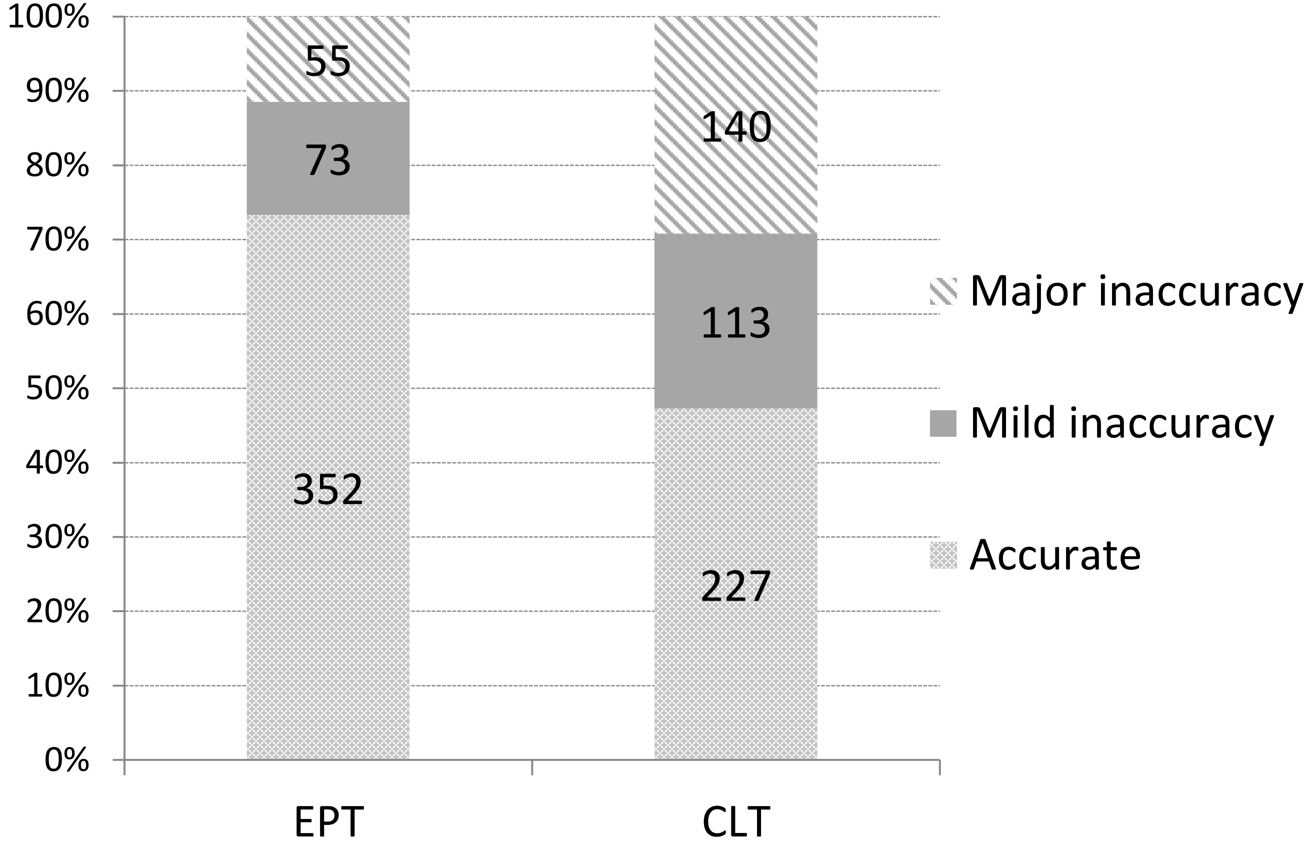


mmHg
150.0
140
130
120
110
100
90
80
70
60
50
40
35
30
25
20
15
10
5
0
-5
-10.0
FS
FS
FS
DCI: 1384.2
5.9 mmHg
IRP = 5.9 mmHg
10s
Swallow #1
4:43.6 min
150.0
27.6
0.0
150.0
11.9
0.0
150.0
68.5
9.1
150.0
40.0
6.5
150.0
-1.6
-1.2
50.0
3.0
0.0
eSleeve
5.4
50.0
3.9
0.0
Gastric
Residual LES pressure = 1.5 mmHg
10s



mmHg
150.0
-10.0
U-3
UES
L-11
L-7
LES
GST
FS
DCI: 0.1
20.0 mmHg
IRP = 20.0 mmHg
10s
Swallow #1
1:38.1 min
eSleeve
Gastric
Residual LES pressure = 8.3 mmHg
10s

Figure 3


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
55
73
352
140
113
227
EPT
CLT
Major inaccuracy
Mild inaccuracy
Accurate